

GIVING ILLUSIONISM A RUN FOR ITS MONEY

Emmett L. Holman

eholman@gmu.edu

For some years now there has been a growing feeling that an impasse has been reached between mainstream physicalist and mainstream dualist theories of consciousness. The former, so it is argued, cannot do justice to the subjective or phenomenological nature of consciousness, and the latter must either deny the causal closure of the physical (for which there is abundant evidence) or embrace epiphenomenalism. As a result, a number of non-mainstream views have lately been getting increasing attention. Among these is the view known as ‘illusionism’; the view that conscious experience—or at least the kind of conscious experience that seems to undermine physicalism—is an illusion. Thus can physicalism be vindicated. Such a view is, of course, highly counter-intuitive, but it is worth giving it a run for its money, which is what I propose to do in this paper. I will take a statement and defense of this view by Keith Frankish (2016a, 2016b) as representative and, looking at what he and some of his critics have to say, see if such a view can be rendered credible. I will conclude that, though the illusionist can answer many of the objections leveled against it, there is one particular objection that seems to be fatal—namely, that it cannot naturalize intentionality in a phenomenologically convincing way.

I

The illusionist begins by conceding to the non-physicalist that conscious experience involves (in a sense) properties—so-called ‘phenomenal properties’—that cannot be accommodated to a physicalist world view. So up to a point Jackson with his knowledge argument, Chalmers with his conceivability/zombie argument, and Levine with his explanatory gap argument, have made their case. However, since, as the illusionist sees it, conscious experience is illusory, that doesn’t mean that phenomenal properties are

actually instantiated; rather they only seem to be; and physicalism cannot be threatened by properties are not instantiated

An obvious question this gives rise to is whether or not the illusionist is denying that there are mental states that have a phenomenology: Is s/he saying that there is nothing it is like to ‘experience’ (or whatever term the illusionist would use) colors or sounds or feel pain? Is s/he saying that we are all zombies, in Chalmers’ sense? On the one hand, it would seem that illusionism is committed to an affirmative answer. Phenomenal properties are supposed to be uninstantiated, according to illusionism, and what can phenomenology consist of if not phenomenal properties? One finds these terms—‘phenomenological’, ‘phenomenal’ and their cognates—used more or less interchangeably in the literature. On the other hand, maintaining that there are no mental states with a phenomenology stretches credibility. So this confronts the illusionist with a dilemma: Can s/he make the case that there can be phenomenology without the instantiation of phenomenal properties. And if not, does that constitute a *reductio* of illusionism? Those are the questions I will be addressing in this paper.

Notice that this challenge is pretty much captured by the popular dismissive argument that illusionism is a non-starter, because to concede that conscious states seem to occur is to concede that they do occur, since seeming is itself a conscious state. Or, as it is sometimes put, there is no appearance-reality distinction in the case of consciousness. So a way of stating the challenge is: Is there a sense of ‘seem’ such that conceding that conscious states seem to occur does not commit us to their actually occurring and thus rendering illusionism self-defeating?

With all this in mind, let’s consider what Frankish says about how the illusion of phenomenal consciousness might work. He entertains two possibilities.

(IL1) “...introspection issues directly in dispositions to make phenomenal judgments—judgments about the phenomenal character of particular experiences and about phenomenal consciousness in general” (p. 14)

(IL2) “...introspection generates intermediate representations of sensory states, perhaps of a quasi-perceptual kind, which ground our phenomenal judgments” (p. 14)

Though opinions and intuitions on this vary, I think it is safe to say that there is a consensus, if not quite a unanimous agreement that the scenario described in IL2 involves phenomenology, keeping in mind that we are currently leaving it an open question whether or not phenomenology involves the instantiation of phenomenal properties. There is more disagreement about the status of the IL1 scenario in this regard, with (probably) a unanimously negative verdict when it comes to the dispositional cases. Of course, when we are entertaining something as counter-intuitive as illusionism, who knows where things eventually will go. But we have to start somewhere.

II

Francois Kammerer (2016) undertakes to explain away the intuition that illusionism is self-defeating. As he sees it, this intuition stems from a model of how illusion—including introspective illusion—works, a model that is part of a theory of mind/epistemology that is hard wired into our sensory and cognitive faculties and thus informs the output of those faculties. But this theory could be false, and Kammerer proposes that it is. Thus can the intuition that illusionism is a non-starter be defused.

That introspective experience is in this way theory laden Kammerer calls the “Theoretical Introspective Hypothesis” (TIH). According to TIH, our introspection is informed with a theory that, among other things, gives an account of what characterizes experiential illusion. My rendering of this account will take some terminological liberties, in line with how I have framed the problem I am addressing. Still, I submit that it gives a fair accounting of how Kammerer sees things. So, according to this account, the phenomenological nature (an expression Kammerer doesn’t use in his exposition) of an illusory experience is the same as the phenomenological nature of the corresponding veridical experience. I.e., an illusory experience of X is phenomenologically identical with a veridical experience of X. The difference

is that in the veridical case, X exists (or is instantiated) and is causally responsible for the phenomenological nature, whereas in the illusory case this is not so. This works fine until we get to the case where X is itself an experience which is being targeted by introspection. Then the illusionist is in trouble, because the phenomenology of the veridical experience has to include the phenomenology of X, the experience being targeted. In that case, in order to honor the phenomenological identity between a veridical and a would-be corresponding illusory experience, the latter has to include X as well, which means that the would-be illusory experience can't be illusory after all. So illusionism, on this model of how experiential illusion works, contains a built-in contradiction—hence, the strong intuition that it's absurd; that there can be no appearance-reality distinction in the introspective case.

So if illusionism is to work, we need a different model of how introspective illusion works. Since the problem arises because the introspective experience must include the introspected experience in its phenomenological nature, we need a model for which this is not so. The obvious move is to construe introspective experience as strictly intentional; a view Frankish himself subscribes to. On such a view, phenomenal properties are, to be sure, included in introspective experiences in a sense, but only as intentional objects, not as actually instantiated properties. So the way is open to hold that the relevant introspective intentional states are all illusory, and so that illusionism with respect to phenomenal properties is true.

III

So, how plausible is it that introspective experience is strictly intentional? Some may think it's not very plausible. The phenomenology of such experiences seems very different from that of the paradigm examples of intentional states such as belief, fantasy thinking, and the like. The former seem to involve a *presence* that the latter lack. Indeed, as earlier indicated, some even question whether the latter essentially

involve a phenomenology at all. Hence, we have the widely held view that experience involves qualia and intentionality doesn't.

This intuition underlies a couple of the criticisms leveled against Frankish. So, Jesse Prinz (2016) says that, given that Frankish drops qualia from the picture, he may say the “seemings [of the presence of phenomenal properties] are not experiences. He could insist that they are mere beliefs (or equivalently he could say we don't seem to have qualia; we just believe we have qualia)” (p. 193). So “...the term ‘illusionism’ might be replaced by ‘mistakenism’” (p. 194). However, he says,

A shift from illusionism to mistakenism would foreclose one of the two ways in which illusionism collapses into reductionism, but it takes on a burden that is hard for the theory to bear. Beliefs, I have noted, are coarser grained than sensory experiences. Therefore, it's not clear that our faith in qualia could come from false beliefs. After all, the things that seems to exist seem to have qualities that outstrip beliefs. (p. 194)

Similarly, James Tartaglia (p. 2016) inveighs against “the physicalist illusionist's substitution of bare judgment for the phenomenology as an account of seeming”. (p. 243)

But the right moral of all this isn't that introspective experience isn't intentional; it's that it can't be identified with such paradigm intentional states as belief or judgment. I.e., IL1 given back is Sec. I is not the right model for how introspective illusion works. We have to go with something like IL2. But this still leaves the illusionist with the burden of showing that an intentional state can have the phenomenology we associate with experience (with “sensory states, perhaps of a quasi-perceptual kind” as Frankish puts it) Here's how I propose the illusionist can go about doing so.

First, as indicated, we take introspective experience to be a type of intentionality other than belief, judgment or any of the other conventionally recognized intentional states. Rather, it is *sui generis*. (As an aside, one might consider the same for perceptual experience). Second, picking up on Kammmer's idea that introspective content is informed by a hard wired theory, we hold that the *sui generis* intentional states

in question are the outputs of introspective mechanisms that are modular. (And maybe the same for perceptual mechanisms as well) The idea is that there are innately endowed limitations on the contents of these outputs, their being restricted to certain properties. In the case of introspection these would be the phenomenal properties. (In the case of perception they would probably be the so-called ‘sensible properties’). Furthermore, given a certain set of sensory stimuli, a more or less algorithmic procedure cranks out the corresponding perceptual/introspective outputs, unaffected (synchronically, at least) by the individual’s background beliefs or other higher cognitive functions. This can account for the relative inflexibility, *vis a vis* such background beliefs and higher cognitive functions, of the contents of the experiences as contrasted with that of the more conventionally acknowledged intentional states. In particular, it can account for why the illusion of phenomenal consciousness does not go away even if we are convinced that it is an illusion, and so add something to Kammerer’s explanation of the counter-intuitive nature of illusionism

Notice that if this is going to succeed, we are going to have to say that the mechanism of introspective awareness *per se* does not have a phenomenology. Rather, the phenomenology is constituted entirely by the introspective intentional content; not the mechanism of introspective intentionality itself. If it were otherwise—in particular if the mechanism of introspective intentionality itself were to instantiate phenomenal properties—then, of course, illusionism could not be upheld. This can be a difficult distinction to make, and this difficulty perhaps adds to the intuition that the appearance of conscious experience, *qua* instantiation of phenomenal properties must itself be a conscious experience *qua* instantiation of phenomenal properties, and so to the intuition that illusionism is a non-starter.

But this needs to be qualified, or at least clarified. Derk Pereboom (2011), who is sympathetic to illusionism, has suggested that introspective intentional states *per se* might have a phenomenology as long as they are themselves the targets of higher level introspective states. In that case, the appearance of phenomenal instantiation on the part of the introspective experience would itself be an illusion which is generated by a higher level introspective state which targets it. And this need not lead to an infinite regress,

says Pereboom. It allows that we can be in an n order introspective state without being in an $n + 1$ order introspective state that targets it. In that case, being in the n order state would not be accompanied by the impression that that state itself instantiates phenomenal properties. The only phenomenal properties apparently present would be those that constitute the intentional content of the n order state. But that, the illusionist will maintain can do full justice to the phenomenology (though one may easily be confused about this. See above). In any case, then, phenomenology is fully accounted for in terms of the content of an introspective (and/or perceptual?) intentional state.

Perhaps, then, putting all this together, we have a credible case for an intentionalist account of introspective and perceptual experience. Or perhaps not. If not, then so much the worse for such an intentionalist account. And then, so much the worse for illusionism.

IV

We thus have a number of moves for rendering illusionism less implausible than it may initially seem. But there is one more problem that needs to be addressed. The whole point of this exercise is to make the world safe for physicalism; so to that end, we need to naturalize—meaning, of course, physicalize, intentionality. This is what Frankish sees as the main challenge for his view. Since the content of all the introspective intentional states is supposed to be illusory, one popular strategy for naturalizing intentionality won't work here. This is the strategy according to which intentional content is to be identified with the actually instantiated property that normally (in some sense of 'normally') causes intentional states of that type. But since phenomenal properties are never, according to illusionism, instantiated, this won't do here. So we must turn to something else, presumably a functional role account. Indeed, Frankish more or less says this a number of times. And, as we'll see, Kammerer seems to subscribe to something like this as well. But more on him later.

First, let's consider how well this can stand up to the standard anti-physicalist arguments. Let's consider, e.g., Chalmers' conceivability argument. It is usually stated more or less as follows:

1. For any being instantiating physical properties and phenomenally conscious properties (or having states instantiating phenomenally conscious properties), it is conceivable that there is a being instantiating the same physical properties but lacking any phenomenally conscious properties. (It is conceivable that a conscious physical being have a zombie twin).
2. If 1, then it is metaphysically possible that there is a being instantiating those same physical properties but lacking phenomenally conscious properties.
3. If it is metaphysically possible that there is a being instantiating those physical properties but lacking phenomenally conscious properties, then physicalism is false.
4. Therefore, physicalism is false.

Since a physical duplicate is also functional role duplicate (assuming the functional role is physically implemented, which I will take as given hereafter) this can be used to argue that a functional role account of consciousness is wrong.

From here, the debate has usually focused on whether or not the first two premises are true: Is there an epistemic gap? Does an epistemic gap entail a metaphysical gap? But the illusionist takes a different tack. To see what this is, let's first spell out exactly what the conclusion says. It says:

- 4'. Therefore, phenomenally conscious properties are not reducible to physical or functional properties.

And the illusionist does not deny this. Indeed, s/he embraces it. Call this form of physicalism whose denial the illusionist embraces 'phenomenal conscious physicalism' (pc-physicalism). But there is another form of physicalism—call it 'actual world physicalism' (aw-physicalism)—that holds that there are no actually instantiated properties that are not reducible to physical/functional properties. To get the denial of this from

4' one must affirm that phenomenally conscious properties are instantiated. So the full argument would be completed as follows:

5. There are actual instantiations of phenomenally conscious properties.
6. Therefore, there are actual instantiations of properties not reducible to physical/functional properties (aw-physicalism is false)

But, of course, the illusionist denies step 5. Phenomenally conscious properties may constitute the content of introspective intentional states, but they are not actually instantiated. So the conclusion is blocked.

But this doesn't quite settle the issue. That is because we are leaving open the question of whether or not phenomenal properties exhaust the phenomenology of introspective experience. So the conceivability argument needs to be reframed so as to beg no questions on that score. To that end I will use the expression 'phenomenological state with phenomenal property content', or 'phical-w-ph-cont' to designate the introspective states targeted in the conceivability argument (and other anti-physicalist arguments). And in conformity with my goal, I will use that expression in as non-committal, non-question begging a way as possible.

The one thing my use of this expression does assume is that introspective experience has a phenomenology; that there is something it is like to undergo such experience; that we are not zombies. Of course, as indicated in some earlier remarks, whether or not that is so is itself up for grabs. But for the time being I am assuming that that is so; hence my adoption of IL2 rather than IL1 as the model for how the illusion of consciousness works. Beyond that, however, the use of the expression is intended to be non-committal. It assumes neither that the phenomenology is exhausted by the phenomenal property content nor that it isn't (hence, 'content' is being used very broadly) nor does it assume even that this phenomenological state is intentional, or that it isn't. It's just intended to designate the what-its-like nature that is involved in introspective experience, whatever that may turn out to involve. With that in mind, let's now reframe the conceivability argument:

1a. For any being instantiating physical properties and the property of being a phical-w-ph-cont,, it is conceivable that there is a being instantiating those physical properties but lacking the property of being a phical-w-ph-cont..

2a. If 1a, then it is metaphysically possible that there is a being instantiating those physical properties but lacking the property of being a phical-w-ph-cont.

3a. If it is metaphysically possible that there is a being instantiating physical properties but lacking the property of being a phical-w-ph-cont, then the property of being a phical-w-ph-cont is not reducible to physical or function properties.

4a. Therefore, the property of being a phical-w-ph-cont is not reducible to physical/functional properties. (pc-physicalism is false)

5a. There are instantiations of the property of being a phical-w-ph-cont.

6a. Therefore, there are actual instantiations of properties not reducible to physical/functional properties (aw-physicalism is false).

1a certainly has no less credibility than 1, and it is not part of the illusionist's strategy to question 1. Of course, there may be no difference between them, and the upshot of all this is that, as far as the metaphysical issues are concerned, it doesn't matter. In particular, it doesn't matter if the phenomenology of introspective experience is exhausted by phenomenal properties or not; in either case we are committed to a non-physicalist theory of mind. The same holds for whether or not introspective experience is intentional, and if so whether or not the content is illusory. If it is, we just conclude that the intentional state *per se* has a non-physical phenomenology. So this shows that naturalization, *via* functional role or anything else, won't work. And finally, the popular quick dismissal of illusionism—that seeming to be a conscious state is itself a conscious state—has, in a sense, been vindicated. At the least, we can say that the phenomenological state is non-physical even if it involves more than instantiated phenomenal properties. So what does it matter if 'seeming' to be conscious should be seen as a conscious state or not; it is, in any case, non-physical.

Or so the reframed argument contends. But let's see if the illusionist has an answer.

Let's first see how Kammerer (2016) thinks we must characterize the illusion of consciousness if are to make illusionism work. He says we must adopt:

...not the naïve concept of illusion, provided by our naïve theory of mind/epistemology, but rather a functional concept of illusion. For example, we have to use a concept of illusion that states that an illusion simply consists in, say, a state that grounds an ongoing, systematic, and cognitively impenetrable disposition to believe something false. (p. 136)

This is rather compactly stated, with a number of questions left unanswered, but here is what I take away from it. First, the functional process he describes has a life of its own, its sole purpose, as it were, being to crank out the beliefs in question irrespective of their truth value. To put it a little less metaphorically, even if there were phenomenal properties, the beliefs in them generated by this process would not count as knowledge. They would be true, to be sure, but only accidentally so. Their truth would not be appropriately linked to the facts that make them true. Correspondingly, if we want to construe the 'grounding state' referred to in this passage as an experience (but see below) that experience would not count as veridical.

But that last remark brings us back to our non-physicalist argument. By an 'experience' we have to a phical-w-ph-cont and our argument purports to show that this can't be physical and hence that there are creatures that instantiate non-physical states. How might Kammerer, or any illusionist, answer this? I don't see any move available other than to reject 5a and hold that the illusory introspective state does not have a phenomenology, which means it is not a phical-w-ph-cont, or what Frankish calls a 'quasi-perceptual state'. So we have to drop IL2 as the right account of how the illusion works and go with IL1 instead. And we have to hold that if there is a 'grounding state' as Kammerer puts it, it is not an experience, not a phical-w-ph-cont, but something like a cause.

Of course, this represents progress for the illusionist only if belief, or judgment, is also not a phical-w-ph-cont ; only if it does not involve a phenomenology; only if there is not something it is like to entertain

a belief or make a judgment. Now there is a school of thought according to which this is so. It is, to be sure, a contested school of thought, especially with the recent rise of the phenomenal intentionality research program. But it is a school of thought to which the illusionist if going to have to adhere.

So here is how the illusionist is going to have to play it: The illusion of phenomenal consciousness arises by introspection's generating false beliefs in the instantiation of non-physical phenomenal properties (or, if one wishes, phenomenological properties with phenomenal content). This occurs because the introspective mechanism is informed by a theory that is hard wired and impenetrable by higher order cognitive faculties. Included in this theory is the proposition that there is a phenomenology underlying all this, one on which our beliefs in phenomenal properties is based, and hence that IL2 gives the right account of how phenomenal beliefs arise. But this theory is false. There is no such phenomenology; hence IL1 is the accurate way of looking at it. So it looks like the illusionist must say we are zombies even though we believe we aren't. There really is nothing it's like to introspect phenomenal properties, even though we believe there is.

This is a hard pill to swallow; and Frankish, for one, tries to resist it. But what is wrong with the foregoing line of reasoning? Well, Frankish alludes to what he sees as a fundamental confusion between instantiating phenomenal properties and having a phenomenology. So, he says "[illusionists] will deny the equivalence between having an inner life and having phenomenal consciousness" (p. 23). But then this seems to resurrect the problem raised by our final, reframed argument. If we are not going to reject 5a of that argument, aren't we back to a non-physicalist conclusion? With that question in mind, let's consider the following passage, the first part of which reiterates the bit about a confusion, but which then goes on to suggest how Frankish would resist the anti-physicalist argument:

But aren't phenomenal properties precisely what makes experience *like something*? That is certainly a common way of construing what-it's-like talk, but there is another way. Illusionists can say that one's experiences are like something if one is aware of them in a functional sense, courtesy of introspective representational mechanisms. (p. 23)

Picking up on this, the following passage is worth quoting:

It may be objected that we can imagine a creature representing itself as having phenomenal properties while still lacking an inner life. Zombies believe they are phenomenally conscious (in some sense at least, arguably they lack full-blown phenomenal concepts ...). But—it may be said—this does not give them an inner life like ours. I am not sure this is obvious. (p. 23)

Elaborating on the idea that it is not obvious, he goes on to say of the illusion of phenomenal properties:

...it depends on a complex array of introspectable sensory states, which trigger a host of cognitive, motivational and affective reactions. If we knew everything about these states, their effects, and our introspective access to them, then, illusionists say, we could not clearly imagine a creature possessing them without having an inner life like ours.

So it appears that Frankish doesn't reject 5a but 1a.

Well, one could take this line, but once one does, one wonders what advantage illusionism enjoys over straightforward old time (as it were) non-illusionist Type A physicalism. For non-illusionist Type A physicalism, the challenge is to show that phenomenal properties are reducible to physical/functional properties and that premise 1 of the original argument is wrong. For illusionist Type A physicalism (which is what Frankish is offering) the challenge is to show that phenomenology is reducible to physical/functional properties and that premise 1a of the reformulated argument is wrong. I don't see that there is much to choose between them.

So if illusionism is to distinguish itself from non-illusionist Type A physicalism, I think it is going to have to take the position that the illusion amounts to introspection generating false beliefs directly rather than false beliefs' being based on quasi-perceptual states that are illusory (IL1 rather than IL2), beliefs having no phenomenology, and in general our mental states' having no phenomenology. In other words, we are zombies. But part of the illusion consists in the fact that we don't believe this. Rather we believe

that there is phenomenology, and given the fact that the disposition to so believe is hard wired into us, illusionism comes across as *prima facie* absurd.

So the most counter-intuitive version of illusionism is the only one that has a chance of succeeding. Some may see this as a *reduction* of illusionism; but of course, some don't. Most notable, no doubt, is Daniel Dennett, who has made a career out of vigorously defending such a view. But I won't comment on that here.

REFERENCES

- Bruntrop, B. and Jaskolla, L. (ed.) (2017): *Panpsychism: Contemporary Perspectives*, Oxford, Oxford University Press.
- Chalmers, D. (2006): "Perception and the Fall from Eden" in Gendler and Hawthorne (ed.).
- Chalmers, D. (2017a): "Panpsychism and Panprotopsychism: in Bruntrop and Jaskolla (ed.).
- Chalmers, D. (2017b): "The Combination Problem for Panpsychism" in Bruntrop and Jaskolla (ed.).
- Coleman, S. (2012): "Mental Chemistry: Combination for Panpsychists, *Dialectica*, vol. 66, pp. 137-166.
- Coleman, S. (2017): "Panpsychism and Neutral Monism: How to Make Up One's Mind" in Bruntrop and Jaskolla (ed.).
- Frankish, K. (2016): "Illusionism as a Theory of Consciousness", *Journal of Consciousness Studies*, vol. 23, #11-12, pp. 11-39.
- Gendler, T. S. and Hawthorne, J. (ed.), (2006): *Perceptual Experience*, Oxford and New York, Oxford University Press.

Kammerer, F. (2016): “The Hardest Aspect of the Illusion Problem—and How to Solve It”, *Journal of Consciousness Studies*, vol. 23, #11-12, pp. 124-139.

Kammerer, F. (2018): “Can You Believe It? Illusionism and the Illusion Meta-Problem”, *Philosophical Psychology*, vol. 31, #1, pp. 44-67.

Kriegel, Uriah (ed.) (2013a): *Phenomenal Intentionality*, Oxford and New York, Oxford University Press.

Kriegel, Uriah (2013b): “The Phenomenal Intentionality Research Program” in Kriegel (ed.).

Pereboom, D. (2011): *Consciousness and the Prospects of Physicalism*, Oxford and New York, Oxford University Press.

Pereboom, D. (2016): “Illusionism and Anti-Functionalism about Phenomenal Consciousness”, *Journal of Consciousness Studies*, vol. 23, #11-12, pp. 172-185.

Prinz, J. (2016): “Against Illusionism”, *Journal of Consciousness Studies*, vol. 23, #11-12, pp. 186-196.

Tartaglia, J. (2016): “What Is at Stake in Illusionism?”, *Journal of Consciousness Studies*, vol. 23, #11-12, pp. 236-255.